NONLINEAR MODELS FOR THE STATISTICS OF ADAPTIVE WAVELET PACKET COEFFICIENTS OF TEXTURE

Johan Aubray[†]*, Ian H. Jermyn, and Josiane Zerubia*

Ariana, joint research group CNRS/INRIA/UNSA
INRIA, 2004 route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex, France
phone: +33 (0)4 9238 7857, fax: +33 (0)4 9238 7643, email: firstname.lastname@sophia.inria.fr
web: www-sop.inria.fr/ariana/

ABSTRACT

Probabilistic adaptive wavelet packet models of texture provide new insight into texture structure and statistics by focusing the analysis on significant structure in frequency space. In very adapted subbands, they have revealed new bimodal statistics, corresponding to the structure inherent to a texture, and strong dependencies between such bimodal subbands, related to phase coherence in a texture. Existing models can capture the former but not the latter. As a first step towards modelling the joint statistics, and in order to simplify earlier approaches, we introduce a new parametric family of models capable of modelling both bimodal and unimodal subbands, and of being generalized to capture the joint statistics. We show how to compute MAP estimates for the adaptive basis and model parameters, and apply the models to Brodatz textures to illustrate their performance.

1. INTRODUCTION

The ability to describe, analyse, and classify textures is an integral part of image processing due to the frequent association of particular textures with particular entities in the scene. For example, in remote sensing, which is our target application, satellite sensor resolutions of the order of a metre mean that some entities (forest, fields,...) can be characterized by their texture, something that was not possible with lower resolutions. The result is a large body of work on texture analysis, modelling, and classification. Despite this attention, however, it cannot be said that texture is fully understood. For example, as will be discussed later, many probabilistic models of texture in the literature describe ensembles in which the most probable images are constant, rather than textured.

In [4], Brady *et al.* studied probabilistic adaptive wavelet packet models of texture. The notion of adapting wavelet packet bases to the data was introduced at the same time as wavelet packets themselves [7], and the idea was swiftly applied to texture classification [5, 14]. Brady *et al.* embedded the idea in a Bayesian probabilistic framework in which the adapted basis and model parameters were simultaneously estimated. This justifies the criterion used for adaptivity, makes underlying assumptions explicit, and provides as output an image model that can be used for any subsequent purpose. Probabilistic adaptive wavelet packet models are parameterized by a dyadic division of (one quadrant of) the Fourier domain (given a mother wavelet, this defines a wavelet packet basis) and, for each subband, a set of parameters defining the probability distribution for the coefficients in that subband. (The mother wavelet can also be included as a parameter, and estimated [1].) This type of model presents several advantages for texture modelling. All wavelet approaches [15] have the advantage that they are easily adapted to regions of different shapes, at least approximately, unlike the Fourier basis. However, like the pixel basis, the assumption of independence between all wavelet coefficients is not sufficient to capture the structure present in textures. In particular, the standard wavelet basis only captures long range dependencies at low frequency, whereas typical dependencies in textures exist at medium frequencies also. In the pixel basis, this problem is addressed using Markov random fields, and there is a great deal of work on using such distributions to model texture [9]. However, the range of the dependencies present in many textures make such an approach computationally difficult. Using wavelet bases, Markov [3] and hidden Markov tree [6] models have been used to introduce dependencies between different scales. Originally designed to capture the structure of edges in the image, these models are not particularly well adapted to the type of within-scale dependencies necessary to represent strong periodicities, for example. Such within-scale dependencies turn each level of the wavelet decomposition into a Markov random field, thus again making the problem computationally hard. Adaptive wavelet packet bases capture long-range dependencies while maintaining computational tractability by incorporating the dependencies into the structure of the basis itself, while maintaining independence between coefficients. Highly adapted subbands effectively link together nearby wavelets, thus incorporating dependencies, while the fall-off of dependence with distance means that the wavelet packet coefficients can more reasonably be supposed to be independent. This is similar to the approach taken with 'geometric' wavelets (curvelets, ridgelets, etc.), where a different basis is used to capture the dependencies implied, for example, by an edge.

In [4], Brady *et al.* used Gaussian distributions for the subbands. They discovered, however, that in the subbands most closely adapted to the texture, the statistics, far from being Gaussian, or even generalized Gaussian, were bimodal. *A posteriori*, the reason for this is clear. The standard leptokurtic statistics for wavelet coefficients arise because the wavelet filters respond weakly in smooth parts of an image, responding strongly only at high image gradients. In the case of a wavelet packet basis adapted to a texture, however, the most highly adapted wavelet filters respond strongly almost everywhere, giving few zeros and many reasonably large values. The bimodal statistics mean that samples from the adap-

[†]The work of the first author was partially supported by the Direction Générale de l'Aviation Civile.

tive wavelet packet histograms are very likely to contain non-zero values of the wavelet packet coefficients in bimodal subbands. This is as it should be: models of texture should describe textured images. This observation points out a problem with many texture models in the literature: because the subband distributions are peaked at zero, the most probable images under these distributions are nearly constant, rather than textured. The different widths of these distributions may enable one texture to be distinguished from another, but the resolving power of a model is greatly increased when it is realized that for some subbands, there are likely to be very few zero coefficients; these are the coefficients that characterize the structure of the texture. This allows bimodal subbands to be used on their own to classify textures, corresponding to a large reduction in description length.

The bimodal statistics perform a function similar to the deterministic component in the Wold decomposition [11]. There, the almost-certain presence of certain frequencies is indicated by a delta function in the (diagonal) covariance: the variance of these Fourier components is infinite. However, very large variance is not required for deterministic presence. Indeed it has the disadvantage that the entropy of the distribution is thereby greatly increased, thus decreasing the accuracy of the description. Bimodal statistics represent the same behaviour by assigning a much higher probability to a finite value than to zero, and they do this without drastically increasing the entropy of the distribution with respect to other subbands.

In [8], Cossu *et al.* introduced models for these new bimodal statistics based on mixtures of Gaussians. These models are very effective, for example, allowing successful segmentations based on single bimodal subbands. They suffer from some drawbacks however. First, they require a choice of model for each subband: generalized Gaussian or Gaussian mixture. This introduces some arbitrariness into the parameter estimation procedure. The second drawback relates to the joint statistics of wavelet subbands. Joint statistics are essential for accurate texture description because they are related to correlations between Fourier components, which define the structure of the texture, as opposed to its main periodicities. Cossu *et al.* studied the joint statistics of bimodal subbands, and discovered that in many cases the subbands were far from independent, instead exhibiting strong and organized dependencies. An example is shown in figure 1, which shows a scatterplot of pairs $(w_{\alpha,i}, w_{\alpha',i})$ of undecimated wavelet packet coefficients from two subbands, bimodal for one texture (dark grey or red points) and unimodal for another (light grey or green points). Notice how well separated the clusters are. Modelling these dependencies using the mixture of Gaussian model developed for the one-point statistics is unwieldy, and so a different approach is needed.

We begin this task by introducing a new model for the bimodal subbands. Instead of a model choice for each subband, a parametric model that interpolates between bimodal and Gaussian statistics is used. In this paper, we use this model to capture the one-point statistics of subbands, but the utility of the approach is that the same family of models can be used to capture the observed joint statistics also. The family has other advantages too; we discuss these further in the conclusion.

In the next section, section 2, we define the parametric model that we will use. In section 3, we discuss the parameter estimation problem and how we solve it. Then in section 4, we present the results of applying the new model to various textures. Finally, in section 5, we conclude and discuss how the models can be extended to capture joint statistics and other properties of textures.

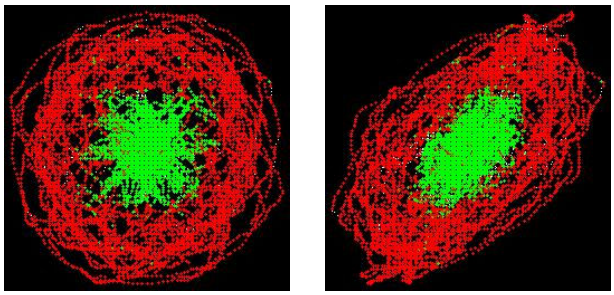

Figure 1: Scatterplot of adaptive wavelet packet coefficients from two bimodal subbands, for the texture on which the basis was trained (red points) and for another texture (green points).

2. MODEL DEFINITION

Adaptive wavelet packet models take the following form:

$$P(I|\theta,T) = \prod_{\alpha\in T} P(w_\alpha|\theta_\alpha,\alpha)\,, \tag{2.1}$$

where: $I:\Omega\to\mathbb{R}$ is the image on some domain Ω; T is a dyadic partition of one quadrant of the Fourier domain; α labels the elements of the partition, the subbands; $\theta: T\to\mathbb{R}^n$ gives the parameters for each subband α; and w_α are the wavelet packet coefficients of I in subband α. We will further assume that the coefficients in each subband are independent:

$$P(w_\alpha|\theta_\alpha,\alpha) = \prod_{i\in\alpha} P(w_{\alpha,i}|\theta_\alpha,\alpha)\,. \tag{2.2}$$

To define the model, it therefore suffices to give the distribution for a single wavelet packet coefficient in each subband. In order to model the bimodal statistics of the wavelet packet coefficients of textures, we choose the following parametric family:

$$P(w_{\alpha,i}|f_\alpha,g_\alpha) \propto \exp\left\{-\left(f_\alpha w_{\alpha,i}^2 + g_\alpha w_{\alpha,i}^4\right)\right\}\,, \tag{2.3}$$

where $\theta_\alpha = (f_\alpha, g_\alpha)$ are the parameters associated with subband α.

For $g_\alpha = 0$, the distribution is Gaussian, and on using equations (2.1) and (2.2), we recover the family of adaptive wavelet packet basis distributions used by Brady *et al.* in [4]. When $f_\alpha, g_\alpha > 0$, the distribution has a flattened form. More interestingly, when $g_\alpha > 0$ and $f_\alpha < 0$, the distribution is bimodal. Thus different parameters in different subbands allow us to interpolate between these types of distribution. Figure 2 shows some examples.

3. PARAMETER ESTIMATION

In order to use the models, we must be able to estimate their parameters given sample data corresponding to the class of

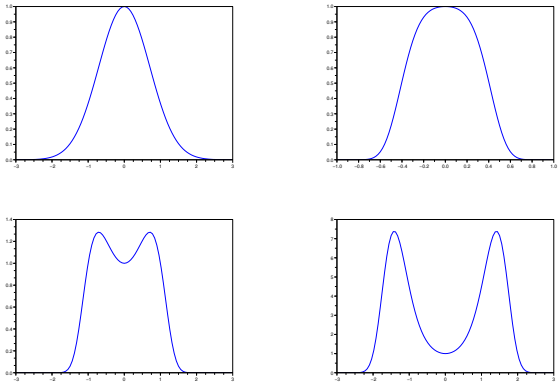

Figure 2: Equation (2.3). Top left: $f = 1$, $g = 0$; Top right: $f = 1$, $g = 20$; Bottom left: $f = -1$, $g = 1$; Bottom right: $f = -2$, $g = 0.5$.

textures we wish to model, for example, images of forest. The posterior probability for the parameters f, g, and T is

$$P(f,g,T|I,K) \propto P(I|f,g,T)P(f,g|T,K)P(T|K) ,$$

where we use K to indicate any other prior knowledge or parameters that may exist. We will take $P(f,g|T,K)$ to be Jeffrey's prior. Invariant MAP estimation [13] will then result in this prior being cancelled by an underlying density, and we can ignore it. We will take $P(T|K) \propto \exp -\beta|T|$, where $|T|$ is the number of elements in the partition T. This penalizes complex partitions. The forms of $P(I|f,g,T)$ and $P(T|K)$ both factorize over the elements of the partition, meaning that we can estimate the parameters using the standard depth-first search procedure over the tree of partitions used in [4, 8]. The decision as to whether to decompose a given node of the tree depends only on the data and parameters for the corresponding subband—the parameters in turn depending only on the data in the subband—and the optimal choices for the child nodes. Thus, on returning to the root of the tree, the optimal partition and corresponding parameters are known. The only missing step is the estimation of the subband parameters from the data in a given subband.

In order to estimate these parameters, we first need to know the normalization constant $Z(f,g)$ for equation (2.3). This takes different forms for different ranges of f and g. The expressions are shown in equation (3.1) (we define $h = f^2/8g$):

$$Z(f,g) = \begin{cases} \frac{1}{2}\sqrt{\frac{-f}{g}}e^h\left(\sqrt{2}\pi I_{-\frac{1}{4}}(h) - K_{\frac{1}{4}}(h)\right) & f<0, g>0 , \\ \frac{1}{2}\sqrt{\frac{f}{g}}e^h K_{\frac{1}{4}}(h) & f>0, g>0 , \\ \sqrt{\frac{\pi}{f}} & f>0, g=0 , \\ \frac{\pi}{\sqrt{2}\Gamma(\frac{3}{4})g^{\frac{1}{4}}} & f=0, g>0 , \end{cases} \tag{3.1}$$

where I and K are modified Bessel functions of the first and second kinds respectively. To avoid numerical problems when $h \gg 1$, we use an expansion in powers of g when h is large:

$$\int_{\infty}^{\infty} dx\, e^{-fx^2-gx^4} \simeq \int_{\infty}^{\infty} dx\, e^{-fx^2}(1 - gx^4 + \frac{1}{2}g^2x^8 + \cdots) .$$

This results in the following approximate expression for Z:

$$Z(f,g) = \sqrt{\frac{\pi}{f}}\left(1 - \frac{3g}{4f^2} + \frac{105g^2}{32f^4} + \cdots\right) .$$

The optimization problem involves only the log likelihood L_α for each subband (we drop the subscript α for ease of reading):

$$L(w|f,g) = -n\ln Z(f,g) - f\sum_i w_i^2 - g\sum_i w_i^4 .$$

where n is the number of coefficients in the subband. The optimization problem is then to maximize L_α over f and g given the constraint $g \geq 0$. Thus we wish to solve the equations

$$L_f = 0 = -n\frac{Z_f}{Z} - \sum_i w_i^2$$
$$L_g = 0 = -n\frac{Z_g}{Z} - \sum_i w_i^4 ,$$

where subscripts indicate differentiation. To perform this optimization, we use a quasi-Newton algorithm [12] with the DFP updating method [10]. Standard Newton-Raphson optimization does not work because it frequently overshoots the barrier $g = 0$, creating negative values of g and an undefined normalization constant. The quasi-Newton algorithm takes the following steps to find a zero of a function $F : \mathbb{R}^n \to \mathbb{R}^n$ (note that the derivatives of L form a function from $\mathbb{R}^2$ to itself):

- Choose an initial point $x_0 \in \mathbb{R}^n$ and a positive semidefinite matrix H_0 (typically, $H_0 = I$, the identity matrix);
- At step k:
  1. Define a direction $d_k = -H_k F(x_k)$;
  2. Find $t_k > 0$ such that $F(x_k + t_k d_k) < F(x_k)$;
  3. Set $x_{k+1} = x_k + t_k d_k$;
  4. If $F(x_{k+1}) \neq 0$, compute H_{k+1}.

The DFP update method takes the following form (where $\delta x_k = x_{k+1} - x_k$ and $\delta F_k = F(x_{k+1}) - F(x_k)$):

$$H_{k+1} = H_k + \frac{\delta x_k \delta x_k^T}{\delta x_k^T \delta F_k} - \frac{(H_k \delta F_k)(H_k \delta F_k)^T}{\delta F_k^T H_k \delta F_k} ,$$

where superscript T denotes transpose. For further explanation, we refer the reader to the references cited above.

We note that the kurtosis of the distribution in equation (2.3) is always less than 3, which is the kurtosis of a Gaussian distribution. If the sample kurtosis of the coefficients in the subband is greater than 3, we avoid the above optimization and simply set $g = 0$, leaving the trivial problem of estimating the variance of a Gaussian distribution.

The choice of initialization point is important. We use a rather naive but in practice effective device: we set f_0 and g_0 equal to their estimated values if the other were set to zero:

$$f_0 = \frac{n}{2\sum_i w_i^2} \qquad g_0 = \frac{n}{4\sum_i w_i^4} \tag{3.2}$$

This leaves open the sign of f. To set this, we note that the kurtosis of the distribution in equation (2.3) is a monotonically increasing function of f. The threshold value at $f = 0$ is easily computed to be

$$\kappa^* = \frac{\Gamma(\frac{5}{4})\Gamma(\frac{1}{4})}{\Gamma^2(\frac{3}{4})} \simeq 2.18 \, .$$

We set $f \gtrless 0$ if $\kappa(w) \gtrless \kappa^*$, where $\kappa(w)$ is the sample kurtosis.

4. RESULTS

We present here the results on Brodatz textures. More results are available in [2]. Figure 3 shows the decomposition for the D84 Brodatz texture for $\beta = 220$. Frequency increases from top to bottom and from left to right. The black subbands are Gaussian ($g = 0$), the grey subbands have $f, g > 0$, while the white subbands are bimodal ($f < 0$, $g > 0$).

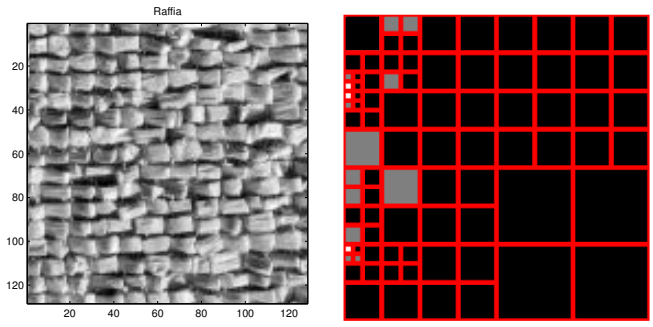

Figure 3: The D84 texture, and its adaptive wavelet decomposition for β = 220. Black indicates a Gaussian subband, gray $f, g > 0$, white a bimodal subband.

Figure 4 shows the histograms of the adaptive wavelet packet coefficients taken from two bimodal subbands. The curves display the fitted models.

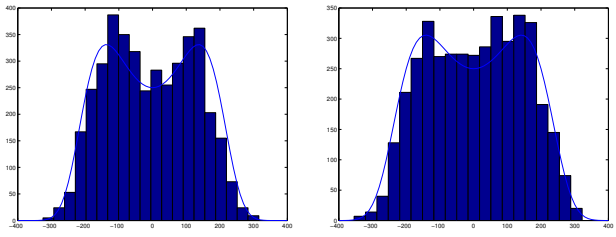

Figure 4: Histograms of the adaptive wavelet packet coefficients of D84 in two bimodal subbands, for $\beta = 220$, together with the fitted models.

Figure 5 shows the decomposition for the D104 Brodatz texture obtained with $\beta = 280$, while figure 6 shows histograms of two bimodal subbands and the fitted models.

5. CONCLUSION

The models described here work well, enabling the diverse observed behaviours of the wavelet packet coefficients to be

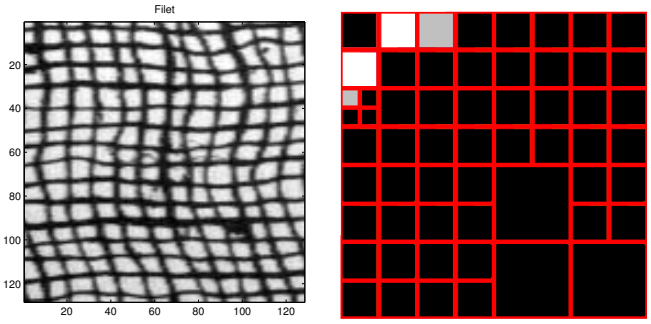

Figure 5: The D104 texture, and its adaptive wavelet decomposition for β = 280. Black indicates a Gaussian subband, gray $f, g > 0$, white a bimodal subband.

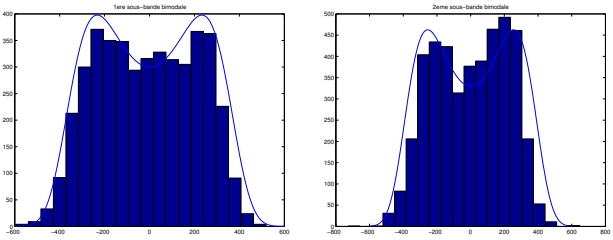

Figure 6: Histograms of the adaptive wavelet packet coefficients of D104 in two bimodal subbands, for $\beta = 280$, together with the fitted models.

captured by a single two parameter model. The resulting image probability distributions can be used, for example, for segmentation. Several points need to be addressed, however, before the potential of the models is fully realized. First, the models are not rotation invariant. While this can be addressed in the usual way using mixture models over rotations via alignment of the texture with the model, this is far from ideal. Intrinsically invariant models would be better. Unfortunately, imposing rotation invariance on a Gaussian distribution leads to a limited class of models that describe unoriented textures. While these have their uses, in many cases they are not sufficient. One of the advantages of the quartic models presented here is that they can be generalized to given rotationally invariant models that nevertheless can describe oriented textures. Second, although capturing the bimodality is a good start in terms of modelling a texture, it does not take into account the relationship between subbands. These dependencies can be complex, as we have seen. Again, the quartic models can be generalized to model these joint statistics. Finally, a further generalization can also capture the phase relationships between Fourier components that are critical to distinguishing certain textures.

References

[1] G. C. K. Abhayaratne, I. H. Jermyn, and J. Zerubia. Texture-adaptive mother wavelet selection for texture analysis. In *Proc. IEEE International Conference on Image Processing (ICIP)*, Genova, Italy, September 2005.

[2] J. Aubray, I. Jermyn, and J. Zerubia. Non-linear models

for the statistics of adaptive wavelet packet coefficients. Research report, INRIA, France, 2006. To appear.

[3] J. S. De Bonet and P. A. Viola. A non-parametric multi-scale statistical model for natural images. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10, Denver, USA, 1998. The MIT Press.

[4] K. Brady, I. H. Jermyn, and J. Zerubia. Texture analysis: An adaptive probabilistic approach. In *Proc. IEEE International Conference on Image Processing (ICIP)*, Barcelona, Spain, September 2003.

[5] T. Chang and C.C.J. Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Trans. Image Processing*, 2(4):429–411, 1993.

[6] H. Choi and R. Baraniuk. Multiscale image segmentation using wavelet-domain hidden markov models. *IEEE Trans. Image Processing*, 10(9):1309–1321, 2001.

[7] R. R. Coifman and M. V. Wickerhauser. Entropy-based methods for best basis selection. *IEEE Trans. Information Theory*, 38(2):713–718, 1992.

[8] R. Cossu, I. H. Jermyn, and J. Zerubia. Texture analysis using probabilistic models of the unimodal and multimodal statistics of adaptive wavelet packet coefficients. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

[9] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.

[10] R. Fletcher. *Practical Methods of Optimization*. Chichester, 2nd edition, 1987.

[11] J. M. Francos, A. Z. Meiri, and B. Porat. A unified texture model based on a 2d Wold-like decomposition. *IEEE Trans. Signal Processing*, 41:2665–2678, 1993.

[12] J.B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms*. Springer-Verlag, 1993.

[13] I. H. Jermyn. Invariant Bayesian estimation on manifolds. *Annals of Statistics*, 33(2):583–605, April 2005.

[14] A. Laine and J. Fan. Texture classification by wavelet packet signatures. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(11):1186–1190, 1993.

[15] S. Livens, P. Scheunders, G. Van de Wouwer, and D. Van Dyck. Wavelets for texture analysis, an overview. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 581–585, Dublin, Ireland, July 1997.